

Free energy calculations of a proton transfer reaction by simulated tempering umbrella sampling first principles molecular dynamics simulations

Yoshiharu Mori*

*Department of Physics, Graduate School of Science,
Nagoya University, Nagoya, Aichi 464-8602, Japan*

Yuko Okamoto†

*Department of Physics, Graduate School of Science,
Nagoya University, Nagoya, Aichi 464-8602, Japan
Structural Biology Research Center, Graduate School of Science,
Nagoya University, Nagoya, Aichi 464-8602, Japan and
Center for Computational Science, Graduate School of Engineering,
Nagoya University, Nagoya, Aichi 464-8603, Japan*

A new simulated tempering method, which is referred to as simulated tempering umbrella sampling, for calculating the free energy of chemical reactions is proposed. First principles molecular dynamics simulations with this simulated tempering were performed in order to study the intramolecular proton transfer reaction of malonaldehyde in aqueous solution. Conformational sampling in reaction coordinate space can be easily enhanced with this method, and the free energy along a reaction coordinate can be calculated accurately. Moreover, the simulated tempering umbrella sampling provides trajectory data more efficiently than the conventional umbrella sampling method.

Understanding chemical reactions by molecular simulations is a challenging problem, because a chemical reaction usually involves bond breaking, which cannot be treated by molecular simulations of classical mechanics based on force fields. We need to use the first principles (or *ab initio*) molecular dynamics (AIMD) methods to deal with bond breaking. However, the time span that can be studied by the first principles molecular simulations is very much limited, because their computational cost is much higher than that of the simulations with force fields.

The molecular simulations based on quantum or classical mechanics are hampered by the multiple-minimum problem, in which simulations tend to get trapped in the local-minimum states of free energy. One can use generalized-ensemble algorithms to overcome this difficulty (for a reviews, see, e.g., [1]). Monte Carlo (MC) and molecular dynamics (MD) simulations based on generalized-ensemble algorithms have been widely performed for many molecular systems in order to have efficient conformational sampling. Three well-known generalized-ensemble algorithms are the multicanonical algorithm (MUCA) [2, 3] (for the MD version see Refs. [4, 5]), the replica-exchange method (REM) [6] (the method is also referred to as parallel tempering [7] and for the MD version, which is referred to as REMD, see Ref. [8]), and simulated tempering (ST) [9, 10].

Recently, general formulations for the multidimensional MUCA, REM, and ST have been given [11–13]. In this article we introduce a special realization of the generalized ST, which we refer to as Simulated Tempering Umbrella Sampling (STUS). This is a generalization of the Umbrella Sampling (US) method [14] and is

closely related to the Replica-Exchange Umbrella Sampling (REUS) in Ref. [15].

Let us consider a system that consists of N atoms, where atom i has coordinate \mathbf{r}_i . We write the set of the coordinates of the atoms as $\mathbf{r} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$. The original potential energy function is represented by $E(\mathbf{r})$, and the temperature by T .

We propose a ST method in parameter space. The parameter here stands for a label that specifies the potential energy function. In this STUS method, M different restraint potential energy functions (or, umbrella potentials) V_1, V_2, \dots, V_M are used. Introducing a reaction coordinate ξ , we define the umbrella potential function V_m by

$$V_m(\xi(\mathbf{r})) = k_m [\xi(\mathbf{r}) - d_m]^2, \quad (m = 1, 2, \dots, M), \quad (1)$$

where k_m are force constants and d_m are equilibrium distances of the reaction coordinate. The STUS simulation yields a uniform probability distribution in parameter (or label) space. It means that a random walk in the M different umbrella potential functions is realized during the simulation.

In the STUS method, each state is specified by coordinate \mathbf{r} and label m , and the following probability distribution function W_m is used:

$$W_m(\mathbf{r}) = \mathcal{Z}^{-1} \exp \{ -\beta [E(\mathbf{r}) + V_m(\mathbf{r})] + a_m \}, \quad (m = 1, 2, \dots, M), \quad (2)$$

where $\beta (= 1/k_B T)$ is the inverse temperature (k_B is the Boltzmann constant), \mathcal{Z} is defined by

$$\mathcal{Z} = \sum_{m=1}^M \int d\mathbf{r} \exp \{ -\beta [E(\mathbf{r}) + V_m(\mathbf{r})] + a_m \}, \quad (3)$$

and a_m are introduced so that the probability distribution in parameter space may be uniform. If the probability in parameter space is constant, a_m are formally written as the following dimensionless free energy except for a constant:

$$a_m = -\ln \left\{ \int dr \exp \{-\beta [E(r) + V_m(r)]\} \right\}. \quad (4)$$

We update the umbrella potential function to another one every few steps during the STUS simulation. When we attempt to change the m -th umbrella potential function V_m to the n -th umbrella potential function V_n , we can use the following transition probability w so that the detailed balance condition may be satisfied:

$$w(m \rightarrow n) = \min \left(1, \frac{W_n}{W_m} \right) = \min (1, \exp(-\Delta)), \quad (5)$$

where

$$\Delta = \beta [V_n(r) - V_m(r)] - (a_n - a_m). \quad (6)$$

We have used the Metropolis criterion [16] to satisfy the detailed balance condition.

The ST parameters a_m ($m = 1, 2, \dots, M$) can be determined by the (multiple-histogram) reweighting techniques applied to a preliminary replica-exchange simulation [11–13]. Once the ST parameters are determined, the STUS simulations are performed by repeating the following two steps: (1) perform a usual molecular simulation with the potential energy function $E + V_m$ for some steps, (2) update the umbrella potential V_m to a

“neighboring” umbrella potential V_n by the transition probability in Eqs. (5) and (6). If accepted, replace V_m by V_n . Go back to step (1).

The free energy as a function of the reaction coordinate ξ , or the potential of mean force (PMF), $\mathcal{F}(\xi_0)$ for the original, unbiased system is defined by

$$\mathcal{F}(\xi_0) = -k_B T \ln \left\{ Z^{-1} \int dr \delta(\xi(r) - \xi_0) \exp[-\beta E(r)] \right\}, \quad (7)$$

where Z is the partition function at temperature T , and δ is the delta function. Note that the umbrella potential functions are not included in this equation.

We can use another expression for PMF instead of Eq. (7):

$$\mathcal{F}(\xi) = -k_B T \ln P(\xi) - C, \quad (8)$$

where $P(\xi)$ is the probability distribution of the reaction coordinate ξ in the original system without umbrella potential functions, and C is an arbitrary constant to set the zero point of the free energy. This form is convenient for molecular simulations because the probability $P(\xi)$ can be obtained by generating a histogram of the reaction coordinate.

From the results of the STUS simulation with the umbrella potential functions, we can calculate $P(\xi)$ using some reweighting techniques such as the multistate Bennett acceptance ratio (MBAR) estimator [17], which is based on the equations in Refs. [18] and [19].

The MBAR equations for calculating the expectation value $\langle A \rangle$ of a physical quantity A are written as follows:

$$\langle A \rangle = \sum_{n=1}^M \sum_{k=1}^{N_n} \frac{A(r_n(k)) \exp[f - \beta E(r_n(k))]}{\sum_{m=1}^M N_m \exp\{f_m - \beta[E(r_n(k)) + V_m(r_n(k))]\}}, \quad (9)$$

$$f = -\ln \sum_{n=1}^M \sum_{k=1}^{N_n} \frac{\exp[-\beta E(r_n(k))]}{\sum_{m=1}^M N_m \exp\{f_m - \beta[E(r_n(k)) + V_m(r_n(k))]\}}, \quad (10)$$

$$f_l = -\ln \sum_{n=1}^M \sum_{k=1}^{N_n} \frac{\exp\{-\beta[E(r_n(k)) + V_l(r_n(k))]\}}{\sum_{m=1}^M N_m \exp\{f_m - \beta[E(r_n(k)) + V_m(r_n(k))]\}}, \quad (l = 1, 2, \dots, M), \quad (11)$$

where N_m ($m = 1, 2, \dots, M$) are the total numbers of trajectory data in the simulation with V_m , and $r_n(k)$ are the k -th coordinate data in the trajectory obtained with V_n . We can obtain the probability distribution $P(\xi)$ of ξ by calculating the expectation values of the number of ξ taking the value ξ_i ($i = 1, 2, \dots$), where ξ_i are discretized

values along the reaction coordinate ξ .

As an application of the STUS method, we considered the intramolecular proton transfer reaction of malonaldehyde (see Fig. 1). We performed STUS AIMD simulations of malonaldehyde in order to show that the STUS method is effective in calculating the free energy of chem-

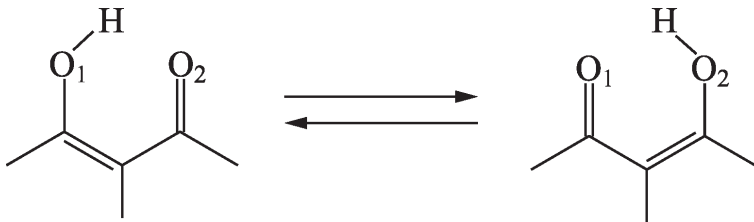


FIG. 1. Intramolecular proton transfer reaction of malonaldehyde. The hydrogen atom that transfers between the two oxygen atoms is written as H. The two oxygen atoms that can bond the hydrogen atom are written as O₁ and O₂.

ical reactions.

We defined the reaction coordinate ξ of the proton transfer as the difference of two distances between the hydrogen atom H and the two oxygen atoms O₁ and O₂:

$$\xi(r) = |\mathbf{r}_{\text{O}_1} - \mathbf{r}_{\text{H}}| - |\mathbf{r}_{\text{O}_2} - \mathbf{r}_{\text{H}}|. \quad (12)$$

When malonaldehyde is in a stable state where hydrogen atom H bonds to oxygen atom O₁, $|\mathbf{r}_{\text{O}_1} - \mathbf{r}_{\text{H}}|$ should be less than $|\mathbf{r}_{\text{O}_2} - \mathbf{r}_{\text{H}}|$ and therefore $\xi < 0$. Likewise, when hydrogen atom H bonds to oxygen atom O₂, $\xi > 0$. When the reaction coordinate has a value which is nearly equal to 0, malonaldehyde should be in a transition state of the proton transfer reaction.

We used 11 different umbrella potential functions in Eq. (1), that is, $M = 11$. k_m were all set to be 0.01 hartree-bohr⁻², and d_1, d_2, \dots, d_{11} were equally spaced between -1.0 \AA and 1.0 \AA , namely, they were set to be -1.0 \AA , -0.8 \AA , \dots , 1.0 \AA , respectively.

We prepared the system of malonaldehyde in 71 water molecules with periodic boundary conditions. We used the CP2K program (version 2.1) [20] to perform AIMD simulations based on the density functional theory with the Born-Oppenheimer approximation. In the density functional calculations, we used the Becke exchange functional [21] and the Lee-Yang-Parr correlation functional [22]. The pseudo potential proposed by Goedecker, Teter, and Hutter [23, 24] was used. A 280 Ry density grid was employed. We used the mixed Gaussian and plane waves approach [25]. We carried out the molecular simulations in the canonical ensemble. The simulation cell was set to be a cubic box ($13.82 \text{ \AA} \times 13.82 \text{ \AA} \times 13.82 \text{ \AA}$). The temperature of the simulation system was set to be 300 K. We used the Nosé-Hoover chain method [26–28] as a constant temperature algorithm, set the number of chains to be 3, and set the time constant of the Nosé-Hoover chain method to be 10 fs. The time step of these simulations was set to be 0.5 fs. The umbrella potential function was changed during the STUS simulation following the transition probability in Eqs. (5) and (6). In the STUS simulation we attempted to update the umbrella potential function to a neighboring one every 10 steps, that is, every 5 fs. We performed four independent simulations with the same conditions described above except for the initial velocities of the atoms. The data were

stored each time just after the ST update attempts. After discarding thermalization steps, we obtained 214 ps simulation data altogether from the four STUS simulations.

In order to show that the STUS method is more efficient than the conventional US method, we also carried out AIMD US simulations with a fixed umbrella potential function using the same 11 umbrella potential functions as in the STUS simulations. Each of the 11 US simulations was performed for 20 ps and the data were stored every 5 fs. Actually, the ST parameters a_m for the above STUS simulations were determined by applying the MBAR reweighting techniques to the results of these 11 US simulations.

The STUS method realizes a random walk in parameter space during the simulation. As a result, the reaction coordinate corresponding to the parameter can be sampled much more widely than in a conventional molecular simulation. Figure 2 shows the time series of the label of the umbrella potential functions and those of the reaction coordinate ξ in one of the four STUS AIMD simulations. Through the simulation, the label of the umbrella functions largely fluctuated between 1 and 11 with almost equal probability. The reaction coordinate also fluctuated greatly, following the change of the umbrella potential. Note that there is an expected strong correlation between the two graphs. The reaction coordinate took on values between around -1.5 \AA and 1.5 \AA . This means that malonaldehyde was able to experience not only the stable states, where hydrogen atom H bonds to either oxygen atom, but also the transition state. Essentially the same results were also obtained for the other three simulations.

From the trajectory data of the STUS AIMD simulations, we calculated the PMF and the probability distribution of the reaction coordinate using the MBAR estimator in Eqs. (9), (10), and (11), which are shown in Fig. 3. The PMF and probability distribution from the results of the conventional AIMD US simulations are also shown in Fig 3.

One would expect that the free energy of the proton transfer reaction of malonaldehyde is a double-well function which has minima at the two stable states and a local maximum at the transition state. Figure 3 (a) shows

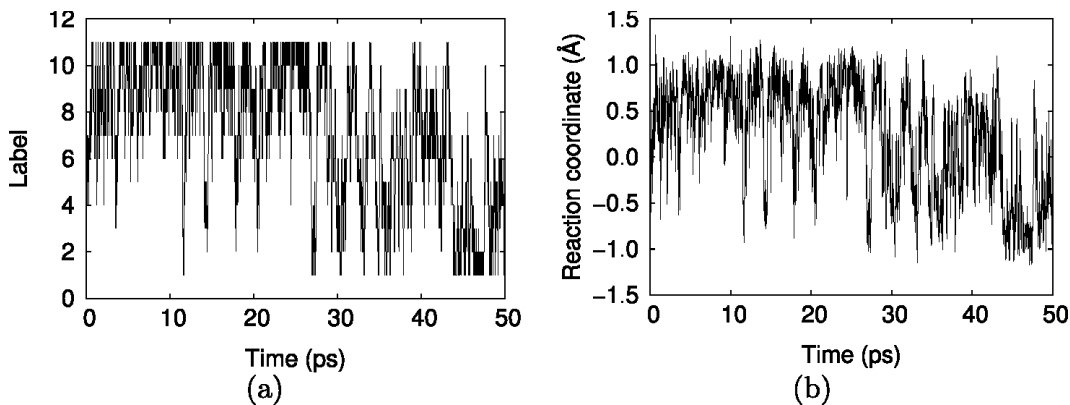


FIG. 2. Time series of (a) the label of the umbrella potential functions and (b) the reaction coordinate ξ during the STUS molecular dynamics simulations.

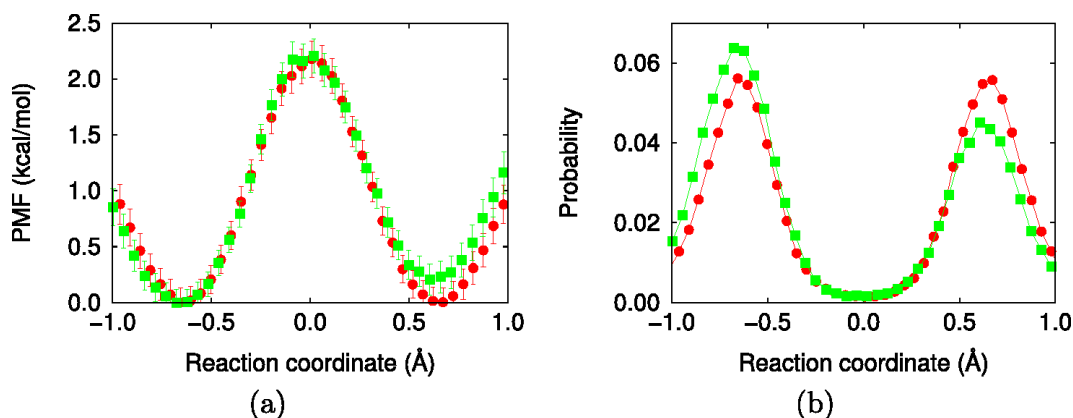


FIG. 3. (color online). (a) Potential of mean force (PMF) of the proton transfer reaction of malonaldehyde. The red filled circles represent the PMF obtained by the STUS simulations and the green filled squares by the conventional US simulations. The error bars were calculated by the MBAR estimator [17]. (b) Probability distribution of the reaction coordinate ξ . The red curve is the probability obtained by the STUS simulations and the green curve is that by the conventional US simulations. The error bars are suppressed to aid the eye.

that the PMF obtained from the STUS simulations has a simple double-well function which has two minima at nearly $\xi = -0.6$ Å and $\xi = 0.6$ Å, and a local maximum at around $\xi = 0.0$ Å.

While the results of the STUS AIMD simulations were able to provide accurate PMF, the PMF calculated from the conventional US AIMD simulations could not be obtained accurately. We can understand this clearly comparing the probability distribution of ξ obtained by the STUS simulations to that by the US simulations. Although the probability in the two stable states is nearly equal to each other in the STUS simulations, the probability obtained by the US simulations is not equally distributed in the two stable states (see Fig. 3 (b)). This implies that the US AIMD simulations could not sample sufficient trajectory data to calculate PMF in the time scale of the present simulations (20 ps for each simulation).

In summary, we have proposed a new simulated tempering method, Simulated Tempering Umbrella Sampling (STUS), and applied it to the first principles molecular dynamics simulations of the intramolecular proton transfer reaction of malonaldehyde. We were able to obtain an accurate potential of mean force of the proton transfer reaction of malonaldehyde from the results of the STUS simulations. We also compared the potential of mean force obtained by the STUS simulations with the one obtained by the conventional US simulations. The STUS method is more efficient in exploring reaction coordinate space than the usual US method.

In the present version of STUS, we fixed the temperature during the simulation. We can easily generalize STUS so that a two-dimensional random walk in both temperature and umbrella potential, as was done in REUS [15].

Moreover, the STUS method can be easily imple-

mented in the existing program packages. One does not need to modify the existing program. Because only the difference of energy is needed in the simulation, one has only to write a simple script that extracts values of energy and reaction coordinates and evaluates the ST transition probability during the simulation.

Some of the computations were performed on the supercomputers at the Information Technology Center, Nagoya University, at the Research Center for Computational Science, Institute for Molecular Science, and at the Supercomputer Center, Institute for Solid State Physics, University of Tokyo. This work was supported, in part, by Grants-in-Aid for JSPS Fellows, for Scientific Research on Innovative Areas (“Fluctuations and Biological Functions”), and for the Computational Materials Science Initiative from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

* E-mail: ymori@tb.phys.nagoya-u.ac.jp

† E-mail: okamoto@phys.nagoya-u.ac.jp

- [1] A. Mitsutake, Y. Sugita, and Y. Okamoto, *Biopolymers* **60**, 96 (2001).
- [2] B. A. Berg and T. Neuhaus, *Phys. Lett. B* **267**, 249 (1991).
- [3] B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).
- [4] U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger, *Chem. Phys. Lett.* **259**, 321 (1996).
- [5] N. Nakajima, H. Nakamura, and A. Kidera, *J. Phys. Chem. B* **101**, 817 (1997).
- [6] K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- [7] E. Marinari, G. Parisi, and J. J. Ruiz-Lorenzo, “Numerical simulations of spin glass systems,” in *Spin Glasses and Random Fields*, edited by A. P. Young (World Scientific, Singapore, 1998) p. 59.
- [8] Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).
- [9] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, *J. Chem. Phys.* **96**, 1776 (1992).
- [10] E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992).
- [11] A. Mitsutake and Y. Okamoto, *Phys. Rev. E* **79**, 047701 (2009).
- [12] A. Mitsutake and Y. Okamoto, *J. Chem. Phys.* **130**, 214105 (2009).
- [13] A. Mitsutake, *J. Chem. Phys.* **131**, 094105 (2009).
- [14] G. M. Torrie and J. P. Valleau, *J. Comput. Phys.* **23**, 187 (1977).
- [15] Y. Sugita, A. Kitao, and Y. Okamoto, *J. Chem. Phys.* **113**, 6042 (2000).
- [16] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [17] M. R. Shirts and J. D. Chodera, *J. Chem. Phys.* **129**, 124105 (2008).
- [18] A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).
- [19] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, *J. Comput. Chem.* **13**, 1011 (1992).
- [20] The CP2K developers group, <http://www.cp2k.org/>.
- [21] A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- [22] C. T. Lee, W. T. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- [23] S. Goedecker, M. Teter, and J. Hutter, *Phys. Rev. B* **54**, 1703 (1996).
- [24] C. Hartwigsen, S. Goedecker, and J. Hutter, *Phys. Rev. B* **58**, 3641 (1998).
- [25] G. Lippert, J. Hutter, and M. Parrinello, *Mol. Phys.* **92**, 477 (1997).
- [26] S. Nosé, *Mol. Phys.* **52**, 255 (1984).
- [27] W. G. Hoover, *Phys. Rev. A* **31**, 1695 (1985).
- [28] G. J. Martyna, M. L. Klein, and M. Tuckerman, *J. Chem. Phys.* **97**, 2635 (1992).